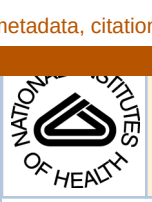

Published in final edited form as:

Biometrics. 2012 March ; 68(1): 12–22. doi:10.1111/j.1541-0420.2011.01650.x.

Multiple Loci Mapping via Model-free Variable Selection

Wei Sun and

Department of Biostatistics, Department of Genetics, University of North Carolina, Chapel Hill, NC, U.S.A

Lexin Li

Department of Statistics, North Carolina State University, Raleigh, NC, U.S.A

Wei Sun: wsun@bios.unc.edu; Lexin Li: li@stat.ncsu.edu

Summary

Despite recent flourish of proposals on variable selection, genome-wide multiple loci mapping remains to be challenging. The majority of existing variable selection methods impose a model, and often the homoscedastic linear model, prior to selection. However, the true association between the phenotypical trait and the genetic markers is rarely known *a priori*, and the presence of epistatic interactions makes the association more complex than a linear relation. Model-free variable selection offers a useful alternative in this context, but the fact that the number of markers p often far exceeds the number of experimental units n renders all the existing model-free solutions that require $n > p$ inapplicable. In this article, we examine a number of model-free variable selection methods for small-n-large-p regressions in the context of genome-wide multiple loci mapping. We propose and advocate a multivariate group-wise adaptive penalization (mGAP) solution, which requires no model pre-specification and thus works for complex trait-marker association, and handles one variable at a time so that works for $n < p$. Effectiveness of the new method is demonstrated through both intensive simulations and a comprehensive real data analysis across 6,100 gene expression traits.

Keywords

Adaptive Lasso; Epistatic interaction; Grouped Lasso; Model-free variable selection; Multiple loci mapping; Sliced inverse regression; Iterative Adaptive Lasso; Multivariate group-wise adaptive penalization

1. Introduction

Modern technologies routinely produce massive amounts of data, and such data deluge now engulfs every branch of science and public life. A typical example is multiple loci mapping in genetics, where a few loci affect the variation of some biological trait, while one has to search over hundreds of thousands of candidate loci on the entire genome. A striking feature of this type of problems is that the number of covariates p often far exceeds the sample size n. In a typical multiple loci mapping problem, the covariates are genotype or copy number of individual genetic markers, the number of which ranges from thousands to millions, whereas the experimental units typically numbers in hundreds to thousands. This challenging small-n-large-p setup renders many classical statistical methods inapplicable. Moreover, the association between the genetic markers and the biological trait can be complex.

In the context of statistical regression modeling, multiple loci mapping manifests itself as the problem of variable selection. An enormous literature on variable selection has appeared

lately; see, e.g., Breiman (1995), Tibshirani (1996), Fan and Li (2001), Zou (2006), Yuan and Lin (2006), Yuan and Lin (2007), Candés and Tao (2007), Fan and Lv (2008), among many others. Most existing approaches assume that the true underlying model is known up to a finite dimensional parameter, and in most cases a linear homoscedastic model is imposed. However, the true model is rarely known in practice, and the true association is very likely to be more complex than a linear relationship. For instance, the genetic effect of copy number variations may exhibit a piece-wise linear pattern – the trait value is proportional to the copy number within a certain range then remains unchanged when the copy number exceeds a threshold, due to the buffering of other factors. Presence of epistatic interactions gives rise to another commonly seen nonlinear pattern in genetics. Although in principle one may include, say, all two-way interactions of the covariates into a linear model, the resulting number of predictors (in the order of p^2) is staggeringly huge given a very large p initially.

In this article, we propose a number of *model-free* variable selection approaches that do not impose any parametric model *before* variable selection. This characteristic distinguishes our proposals from the majority of model-based variable selection methods in the literature. The methods can handle nonlinear associations, as often encountered in multiple loci mapping studies, and are shown to perform superior than the model-based approaches when the true underlying association is different from the imposed linear model. Our proposals stem from the general framework of sufficient dimension reduction (Cook, 1998), and the model-free variable selection methods developed within that framework (Cook, 2004; Li et al., 2005; Ni et al., 2005, 2008; Zhou and He, 2008; Bondell and Li, 2009). However, in all those previous model-free variable selection works, the sample size n is required to be larger, and usually much larger, than the number of covariates p. By contrast, the focus of this article is model-free variable selection when p far exceeds n.

Specifically, we propose and examine two strategies of model-free selection for small-n-large-p regressions. Let $\Sigma = \mathrm{Cov}(X)$ be the $p \times p$ sample covariance matrix of the p covariates. The first strategy aims to avoid inversion of Σ, which is singular when $n < p$. A ridge estimate and a partial least squares (PLS) estimate are employed for that purpose. Nevertheless, this strategy requires estimation of the matrix Σ, and thus in effect handles all p covariates simultaneously. For this reason, we view it as a *global* strategy. The second strategy, which we view as a *local* strategy, refrains from dealing with all covariates together, but instead conducts selection one variable at a time. Toward that end, a multivariate group-wise adaptive penalization (mGAP) approach is proposed. In mGAP, the coefficient of each covariate is updated sequentially by a coordinate decent algorithm, which shares a similar spirit as the classical forward stepwise variable selection, and the latter is shown to possess competent theoretical and empirical advantages for linear models when $n < p$ (Wang, 2009). We investigate both global and local strategies and compare with the model-based solutions, all in the context of genome-wide multiple loci mapping in experimental cross. We find the local strategy achieves a superior performance among all those solutions.

The rest of the article is organized as follows. Sections 2 presents the ridge regression solution and the partial least squares solution. Section 3 proposes the mGAP solution. An extensive simulation study is carried out in Section 4 to examine the performance of the two proposed strategies, as well as to compare with the existing model-based solutions. Section 5 illustrates the advantages of the new method by a real multiple loci mapping data analysis across 6,100 gene expression traits. Section 6 concludes the paper with a discussion.

2. Ridge and Partial Least Squares Solution

2.1 Simultaneous Dimension Reduction and Variable Selection

We first quickly review the framework of sufficient dimension reduction (SDR) and how variable selection is achieved within this framework.

For a regression of a response Y on a p-dimensional covariate vector X, SDR seeks to replace X with a few of its linear combinations, while preserving full regression information and assuming no parametric models. Without loss of generality, we assume X has been standardized to have mean 0 and variance 1 for each covariate. A central parameter of interest in an SDR inquiry is the minimum subspace $\mathcal{S}$ in IR^p such that Y is independent of X given $P_{\mathcal{S}}X$, where $P_{\mathcal{S}}$ is the projection onto $\mathcal{S}$. Such a space is called the *central subspace*, it uniquely exists under minor conditions (Cook, 1996), and it is denoted as $\mathcal{S}_{Y|X}$. There have been many approaches proposed to estimate $\mathcal{S}_{Y|X}$, most of which can be formulated in a unified representation and are collectively referred as inverse regression estimators (IRE) (Cook and Ni, 2005). Specifically, IRE starts with the construction of a $p \times h$ matrix $\theta = (\theta_1, \ldots, \theta_h)$ satisfying $\mathrm{span}(\theta) = \mathcal{S}_{Y|X}$. It then obtains its sample estimate $\hat{\theta}$ through

$$\min_{\eta,\gamma}\left\{\mathrm{vec}(\widehat{\theta}) - \mathrm{vec}(\eta\gamma)\right\}^{\mathrm{T}}\left\{\mathrm{vec}(\widehat{\theta}) - \mathrm{vec}(\eta\gamma)\right\}, \tag{1}$$

over $\eta \in \mathrm{IR}^{p\times d}$, $\gamma \in \mathrm{IR}^{d\times h}$. Letting $(\hat{\eta}, \hat{\gamma})$ denote the corresponding minimizers, then $\mathrm{span}(\hat{\eta})$ forms an estimate of $\mathcal{S}_{Y|X}$. To ensure $\mathrm{span}(\theta) \subseteq \mathcal{S}_{Y|X}$, it requires the linearity condition that $E(X|\beta^{\mathrm{T}}X)$ is linear in $\beta^{\mathrm{T}}X$, where β denotes a basis of $\mathcal{S}_{Y|X}$. This is usually viewed as a mild condition, is widely imposed in the SDR literature, and is true as p tends to infinity (Hall and Li, 1993). In our multiple loci application, we view this condition approximately true since p is very large. We also note that this condition is imposed on the distribution of X rather than on $Y|X$. For this reason, this family of SDR estimators, as well as the variable selection methods developed within this family, are regarded as model-free.

Simultaneous variable selection within this framework is based upon the key observation that all the irrelevant covariates to regression $Y|X$ have their corresponding rows of the basis of $\mathcal{S}_{Y|X}$ equal to zero and vice versa. This connects selection of variables with estimation of $\mathcal{S}_{Y|X}$. Ni et al. (2005) and Bondell and Li (2009) introduced a $p \times 1$ shrinkage vector ω, and proposed to first obtain the estimates $\hat{\theta}$, $\hat{\eta}$ and $\hat{\gamma}$ from (1) and then minimize over ω,

$$\min_{\omega}\left[\mathrm{vec}(\widehat{\theta}) - \mathrm{vec}\left\{\mathrm{diag}(\omega)\widehat{\eta}\widehat{\gamma}\right\}\right]^{\mathrm{T}}\left[\mathrm{vec}(\widehat{\theta}) - \mathrm{vec}\left\{\mathrm{diag}(\omega)\widehat{\eta}\widehat{\gamma}\right\}\right] + \lambda\sum_{j=1}^{p}|\omega_j|, \tag{2}$$

where λ is a non-negative penalty constant. The term $\sum_{j=1}^{p}|\omega_j|$ in (2) is a Lasso type L_1 penalty. As a result, an increasing λ would force some elements of $\hat{\omega}$ to exactly equal zero. The resulting $\mathrm{diag}(\hat{\omega})\,\hat{\eta}$ is taken as a sparse estimate of the central subspace basis, and zero elements of $\hat{\omega}$ shrink the entire rows of the estimated basis to zero. Consequently, one achieves variable selection by screening out those variables whose corresponding $\hat{\omega}_j$'s are zero. The shrinkage vector ω serves as an extra garrote parameter (Breiman, 1995), and the Lasso penalty in (2) can be equivalently replaced by a nonnegative garrote penalty (Bondell and Li, 2009). For brevity, we refer this solution as garrote selection.

2.2 Ridge and Partial Least Squares Based Garrote Selection

The garrote selection hinges on the construction of θ. There are a class of choices of θ that are based upon the first inverse moment $\varphi(Y) = \Sigma^{-1}E(X|Y)$, which all have roots in the seminal sliced inverse regression (SIR) (Li, 1991). Letting $t_1(Y)$, …, $t_h(Y)$ denote a set of h transformation functions of Y, this class of θ takes a common form $\theta_s = E\{t_s(Y)\varphi(Y)\}$, $s = 1$, …, h, where θ_s denotes the sth column of θ. The original SIR corresponds to the sliced indicator function, $t_s(Y) = 1$ if Y is in slice s and 0 otherwise. Other choices of t_s include the B-spline transformation (Fung et al., 2002), the polynomial transformation $t_s(Y) = Y^s$ (Yin and Cook, 2002), and the covariance estimator $t_s(Y) = Y$ if Y is in slice s and 0 otherwise (Cook and Ni, 2006). Moreover, a number of papers (Li, 1991; Fung et al., 2002; Cook and Ni, 2006) have found that this class of methods are not overly sensitive to the choice of h.

We first note that θ_s can be rewritten as $\theta_s = \Sigma^{-1}\sigma_s$, where $\sigma_s = \text{Cov}\{X, t_s(Y)\}$. Thus, estimation of θ_s requires the estimation of Σ^{-1}. When the number of covariates p exceeds the sample size n, the usual sample estimate of the covariance matrix Σ is singular and is not invertible. As such, the usual estimator of θ_s becomes unavailable. This prevents a direct application of the aforementioned variable selection method to our setup. We next consider two solutions to address this problem.

One solution is to introduce a ridge type regularization, i.e., replacing Σ^{-1} with $(\Sigma + \tau_s I_p)^{-1}$ for a positive ridge constant τ_s. Tuning of the ridge constant can be based on the generalized cross validation criterion as in Li and Yin (2008a). This solution is in spirit similar to Li and Yin (2008b). The difference is that Li and Yin (2008b) focused on SIR only, whereas the solution here applies to a class of first inverse moment based reduction methods.

Alternatively, one may employ the idea of partial least squares (PLS) (Wold, 1975) to handle $n < p$. Define $R_{u_s} = (\sigma_s, \Sigma\sigma_s, \ldots, \Sigma^{u_s-1}\sigma_s) \in \text{IR}^{p\times u_s}$ for some positive integer u_s. Li et al. (2007) showed that, under the linearity condition, $R_{u_s}(R_{u_s}^{\text{T}}\sum R_{u_s})^{-1}R_{u_s}^{\text{T}}\sigma_s \in \mathscr{S}_{Y|X}$. This is indeed the population PLS estimator. The reason that PLS works when $n < p$ is that, the matrix $R_{u_s}^{\text{T}}\sum R_{u_s}$ that needs inversion in PLS is only $u_s \times u_s$, compared to the $p \times p$ matrix Σ in OLS. As long as $u_s < n$, the sample counterpart of this matrix is invertiable. In practice, u_s is often small, and its tuning can be based on the eigenvalues of the matrix $R_{u_s}R_{u_s}^{T}$ or some smoothing based criterion (Li et al., 2007).

In our context of model-free variable selection for $n < p$ regressions, we thus suggest to employ $\theta_s = (\Sigma + \tau_s I_p)^{-1}\sigma_s$ or $\theta_s = R_{u_s}(R_{u_s}^{\text{T}}\sum R_{u_s})^{-1}R_{u_s}^{\text{T}}\sigma_s$ as our initial sample estimate $\hat{\theta}$ for θ. One then follows (2) to achieve variable selection. Some remarks are noteworthy. First, similar to garrote selection (Breiman, 1995), variable selection is achieved through a two-step procedure, i.e., one first carries out a dimension reduction basis estimation of $\hat{\theta}$, $\hat{\eta}$ and $\hat{\gamma}$ in (1), and then selects variables through the estimate $\hat{\omega}$ in (2). As a result, we obtain both a dimension reduction basis estimate $\text{diag}(\omega)\hat{\eta}$, and a selection of relevant variables simultaneously. Second, we view both the ridge and the PLS solutions as global approaches, in the sense that they handle all p covariates simultaneously by involving the $p\times p$ covariance matrix Σ in estimation. When p is really large whereas n is small to moderate, such a global strategy may not be the most effective intuitively. This motivates us to consider the next local strategy that conducts variable selection one variable at a time when p far exceeds n.

3. Multivariate Group-wise Adaptive Penalization

3.1 Variable Selection without Reduction Basis Estimation

If the analysis goal is variable selection *only*, one may simplify the aforementioned two-step procedure to a one-step solution. Note that $\theta_s = \Sigma^{-1}\,\mathrm{Cov}\{X, t_s(Y)\}$; that is, a sample estimate of θ_s can be obtained by OLS estimation of regressing $t_s(Y)$ on X. As such, θ can be estimated by applying multivariate OLS to a set of transformations of the response, $t_1(Y)$, …, $t_h(Y)$, given X. Variable selection can then be achieved by shrinking the entire rows of θ to zero. Motivated by this observation, Ni et al. (2008) proposed a method named multivariate adaptive Lasso (MAL) to couple multivariate OLS with adaptive group Lasso penalty (Zou, 2006; Yuan and Lin, 2006). For sample observations $\{(x_i, y_i), i = 1, \ldots, n\}$, let $\boldsymbol{T}_s = (t_s(y_1), \ldots, t_s(y_n))^{\mathrm{T}}$, $s = 1, \ldots, h$, let $\boldsymbol{T}$ denote the $n \times h$ matrix with $\boldsymbol{T}_s$ as its s-th column, and let $\boldsymbol{X}$ denote the $n \times p$ data matrix with x_i as its i-th row, $i = 1, \ldots, n$. Ni et al. (2008) considered the minimization

$$\min_{B}\left(\sum_{s=1}^{h}\left\|\boldsymbol{T}_s - \boldsymbol{X}\,B_s\right\|^2 + \lambda\sum_{j=1}^{p}\frac{\|b_j\|}{\|\tilde{b}_j\|}\right), \tag{3}$$

over the $p \times h$ matrix $B = (b_{js})$, where B_s denotes its sth column, $s = 1,\ldots, h$, b_j denotes its jth row, $j = 1, \ldots, p$, and $\|b_j\| = \sqrt{\sum_{s=1}^{h} b_{js}^2}$. Let $\tilde{B}$ be the OLS estimate of B, and $\tilde{b}_j$ denotes its jth row. In (3), an adaptive group Lasso penalty (Yuan and Lin, 2006) is placed on each row b_j of B, and as a consequence, an increasing λ would force some rows of B to be completely zero, which in turn achieving variable selection.

When $n < p$, the MAL solution of (3) can *not* be directly applied, since the OLS initial estimate $\tilde{B}$ is not available. Similar to the global strategy in Section 2.2, however, one may substitute in (3) a ridge estimator or a PLS estimator for $\tilde{B}$. We adopt this strategy in our numerical studies so that we can compare MAL with other methods when $n < p$. On the other hand, as we have discussed at the end of Section 2.2, given that p is much larger than n and the signal in multiple loci mapping is generally weak, such a global estimator of B using ridge or PLS may not be accurate, which in turn causes inaccuracy in variable selection. Our simulations in Section 4 would further confirm this intuition.

Alternatively, motivated by both forward regression (Wang, 2009) and an iterative adaptive Lasso (IAL) solution developed by Sun et al. (2010), we next propose an iterative variable selection method that extends MAL but handles variables one at a time rather than all together. Unlike forward regression and IAL which assume the linear model, the new solution is model-free and can handle a variety of models such as the piece-wise linear model and the epistatic models. The new method originated from the MAL solution of Ni et al. (2008), but as we will show next, the extension is far from incremental. We call our proposal the multivariate group-wise adaptive penalization (mGAP) solution.

3.2 Objective Function

We first develop the objective function for mGAP. We start with a modification of (3):

$$\sum_{s=1}^{h}\frac{\left\|\boldsymbol{T}_s - \boldsymbol{X}\,B_s\right\|}{w_s} + \lambda\sum_{j=1}^{p}\frac{\|b_j\|}{\kappa_j}. \tag{4}$$

Compared to (3), we introduce two changes. First, we assign different weights w_s to the residual sum of squares of $\boldsymbol{T}_s$ on $\boldsymbol{X}$. This is based upon the intuition that those transformations $\boldsymbol{T}_s$'s are not equally informative for our purpose of variable selection. Second, we replace the OLS estimates $\hat{b}_j$'s with a set of parameters κ_j's. If $n > p$, the parameters w_s and κ_j can be naturally estimated by $\widehat{w}_s=\widehat{\sigma}_s^2=\left\|\boldsymbol{T}_s-\boldsymbol{X}\,\widehat{B}_s\right\|^2/n$, and $\hat{\kappa}_j = \|\hat{b}_j\|$.

Next we observe that, without any restriction, w_s's and κ_j's in (4) are not estimable, because one can always multiply a constant to them to reduce the objective function (4). We thus add restrictions on the sizes of w_s's and κ_j's by considering a modified objective function of (4):

$$\sum_{s=1}^{h}\frac{\left\|\boldsymbol{T}_s-\boldsymbol{X}\,B_s\right\|}{w_s}+\lambda\sum_{j=1}^{p}\frac{\left\|b_j\right\|}{\kappa_j}+\lambda\sum_{j=1}^{p}\log(\kappa_j)+n\sum_{s=1}^{h}\log(\omega_s). \tag{5}$$

The new restrictions make w_s's and κ_j's identifiable. A closer examination reveals that, for a given B, the solutions of κ_j's that minimize (5) are of the form,

$$\widehat{\kappa}_j=\arg\min_{\kappa_j}\left\{\frac{\left\|b_j\right\|}{\kappa_j}+\log(\kappa_j)\right\}=\left\|b_j\right\|,\qquad j=1,\ldots,p.$$

A drawback of this solution is that, if $\|b_j\| = 0$, $\hat{\kappa}_j = 0$, then an infinite penalty is added to the jth covariate so that it would never enter the model again, even when the estimates of other coefficients change and this covariate may now have a significant effect in reducing the residual sum of squares. Taking this into consideration, we further modify the objective function (5) by adding another tuning parameter τ, and come to its final form:

$$\sum_{s=1}^{h}\frac{\left\|\boldsymbol{T}_s-\boldsymbol{X}\,B_s\right\|^2}{w_s}+\lambda\sum_{j=1}^{p}\frac{\left\|b_j\right\|+\tau}{\kappa_j}+\lambda\sum_{j=1}^{p}\log(\kappa_j)+n\sum_{s=1}^{h}\log(\omega_s). \tag{6}$$

We now propose to minimize (6) over the $p\times h$ matrix B, the $h\times 1$ vector $w = (w_1, \ldots, w_h)^{\mathrm{T}}$, and the $p \times 1$ vector $\kappa = (\kappa_1, \ldots, \kappa_p)^{\mathrm{T}}$, for some tuning parameters $\lambda > 0$ and $\tau > 0$. To see how variable selection is achieved one at a time through (6), we first observe that, in (6), the group Lasso type penalty is placed on each row b_j of B. As a consequence, for some λ, the solution $\hat{b}_j$ will be exactly zero, which in effect leads to variable selection. Next, as we will see in Section 3.3, the update of B is to carry out row by row by cycling through b_j's for $j = 1, \ldots, p$, and as such selection is performed in a one variable at a time fashion. We next develop an alternating optimization algorithm to solve (6) with fixed λ and τ, and then propose a criterion for the tuning of λ and τ.

3.3 Optimization and Tuning

We cycle through the steps of fixing B to solve w and κ, and fixing w and κ to solve B. More specifically, for a given $\hat{B}^{(t)}$ at iteration t, we have,

$$\widehat{\kappa}_j^{(t+1)}=\arg\min_{\kappa_j}\left\{\frac{\left\|\widehat{b}_j^{(t)}\right\|+\tau}{\kappa_j}+\log(\kappa_j)\right\}=\left\|\widehat{b}_j^{(t)}\right\|+\tau,$$

$$\widehat{w}_s^{(t+1)}=\arg\min_{w_s}\left\{\frac{\left\|\boldsymbol{T}_s-\boldsymbol{X}\,\widehat{B}_s^{(t)}\right\|^2}{w_s}+n\log(w_s)\right\}=\frac{\left\|\boldsymbol{T}_s-\boldsymbol{X}\,\widehat{B}_s^{(t)}\right\|^2}{n}.$$

For $\kappa_j^{(t+1)}$, we note that the tuning parameter $\tau > 0$ now gives any covariate with a zero coefficient estimate a large (in the scale of λ/τ) but finite penalty. Moreover, the above solutions of $w_s^{(t+1)}$ and $\kappa_j^{(t+1)}$ are intuitively reasonable, because they resemble the solutions when $n > p$, as discussed at the beginning of Section 3.2. Next, given $\widehat{\kappa}_j^{(t+1)}$ and $\widehat{w}_s^{(t+1)}$, we update B by

$$\widehat{B}^{(t+1)}=\arg\min_{B}\left(\sum_{s=1}^{h}\frac{\left\|\boldsymbol{T}_s-\boldsymbol{X}\,B_s\right\|^2}{\widehat{w}_s^{(t+1)}}+\lambda\sum_{j=1}^{p}\frac{\left\|b_j\right\|}{\widehat{\kappa}_j^{(t+1)}}\right).$$

This optimization problem can be solved numerically by a coordinate descent algorithm, i.e., by updating one row b_j at a time for $j = 1, \ldots, p$. Details are given in the Web Appendix A. By default, the initial values of all the coefficients in $\widehat{B}^{(0)}$ are set to be zero.

In (6), λ and τ are two tuning parameters that control the amount of penalty imposed on B. Next we employ a Bayesian information criterion (BIC) to tune these two parameters. After BIC, we further amend our procedure with a backward filtering.

Specifically, we choose λ and τ that minimize the criterion,

$$\mathrm{BIC}=n\log(|\widehat{\Omega}|)+\nu\times\log(n),$$

where $\hat{\Omega}$ denotes the sample estimate of the $h \times h$ covariance matrix $\mathrm{Cov}(\boldsymbol{T}-\boldsymbol{X}\hat{B})$, $|\hat{\Omega}|$ is its determinant, and ν is the degrees of freedom of the form,

$$\nu=\sum_{j=1}^{p}\left[I(\left\|\widehat{b}_j\right\|>0)\left\{\sum_{s=1}^{h}\frac{\widehat{r}_{js}}{\widehat{r}_{js}^2-(\widehat{r}_{js}-1)\overline{b}_{js}\widehat{b}_{js}/\left\|\widehat{b}_j\right\|^2}\right\}\right],$$

where

$$\overline{b}_{js}=\left\|x_j\right\|^{-2}\sum_{i=1}^{n}x_{ij}\left\{t_s(y_i)-\sum_{k\neq j}x_{ik}b_{ks}\right\},\quad\text{and}\quad\widehat{r}_{js}=1+\frac{\lambda\widehat{w}_s}{2(\left\|\widehat{b}_j\right\|+\tau)\left\|\widehat{b}_j\right\|\,\left\|x_j\right\|^2}.$$

In above evaluations of BIC, $\hat{B} = (b_{js})_{p\times h}$ and $\hat{w}$ denote the estimate of B and w at convergence. This BIC can be derived following a setup of seemingly unrelated regressions (Zellner, 1962), although we do not impose such a model in our solution. The degrees of freedom formula can be obtained following Yuan and Lin (2006); see Web Appendix B.

BIC may be too liberal in the small-n-large-p setting (Chen and Chen, 2008), so we further supplement the BIC selection with a backward filtering step. Given the model selected by BIC, the backward filtering iteratively tests the significance of each selected covariate following the ascending order of their coefficient norms. For each selected covariate, we apply a likelihood ratio test (LRT) to test the null hypothesis that the corresponding h coefficients of this covariate are 0 for all h regressions. Specifically, if the j_t-th covariate has the largest coefficient norm at the t-th step of backward filtering, then LRT = $n \log|\hat{\Omega}_1| - \log|\hat{\Omega}_0|$, where $\hat{\Omega}_1$ and $\hat{\Omega}_0$ are the residual covariance estimates before and after dropping the j_t-th covariate. If the resulting LRT p-value is larger than a given threshold, that covariate is dropped, and the next covariate is tested; otherwise the backward filtering is terminated, and all the remaining covariates are retained. The p-value threshold is set as $0.05/p_E$, where p_E is the effective number of independent tests. A conservative choice is to set $p_E = p$, the total number of tests. In this paper, we estimate p_E by the ratio of permutation p-value over nominal p-value (Sun and Wright, 2010).

4. Simulations

4.1 Simulation Setup and Summary

Sun et al. (2010) proposed iterative adaptive Lasso (IAL) for multiple loci mapping assuming a linear model, and compared IAL with nine existing model-based variable selection methods, including marginal regression, forward regression, forward-backward regression, the composite model space approach (CMSA) (Yi, 2004), the adaptive Lasso (Zou, 2006) (with initial regression coefficients from marginal regression since least square solution is not available when $p > n$), the HyperLasso (Hoggart et al., 2008), the Bayesian t (Yi and Xu, 2008), the Bayesian Lasso (Park and Casella, 2008), and the Bayesian Adaptive Lasso (Sun et al., 2010). It was found that IAL achieves the best performance among all the competitors, when the underlying model is indeed linear. For this reason, in our study, we focus the comparison of our newly proposed selection methods with IAL. We also compare with the modified MAL with ridge or PLS initial estimate, and elastic net (Zou and Hastie, 2005).

We briefly describe the implementation of different methods. For mGAP, we have experimented with different θ via different transformations of Y, as we discussed in Section 2.2. We found the results are similar, whereas the spline transformation shows some edge over other choices. For brevity, we only report here the results with the spline transformation. By default, we employ a simple quadratic spline with one inner knot, which delivers a competent performance in both simulations and real data analysis. The results of ridge and partial least squares based garrote selections are similar and thus we only report the results of PLS based solution and refer it as garrote selection. Similarly for MAL, we only present the results using PLS as initial estimate. The elastic net penalty has the form of $\lambda(\alpha|b_j|_1 + (1-\alpha)\|b_j\|^2)$, where $\lambda > 0$ is the parameter that controls the overall penalty, $|b_j|_1$ and $\|b_j\|$ denote the L_1 and L_2 norms of b_j, respectively. The parameter $0 \le \alpha \le 1$ controls the proportion of L_1 and L_2 penalties, which is often fixed in applications. To fully explore the potential advantage of elastic net, we tune both λ and α. All tunings are based on BIC, followed by the backward filtering. Our BIC for MAL is slightly different from the definition in Ni et al. (2008) since their BIC definition assumes $p < n$ and use $RSS_{ols}/(n - p - 1)$ in the goodness-of-fit part of their BIC, where we replace by $\log(|\hat{\Omega}|)$.

We first consider the linear model setup, then the nonlinear model setting, including piecewise linear pattern, epistatic interaction, and interaction with unobserved sub-groups. In summary, our general finding is that, when the true underlying model is linear, mGAP that assumes no model retains a comparable performance as IAL that assumes a linear model. On the other hand, when the true model is nonlinear, mGAP has a superior performance than IAL, elastic net, and MAL. Moreover, in our simulation setup that mimics the real multiple loci mapping where the signal is relatively weak and p far exceeds n, the mGAP solution that adopts a one variable at a time local strategy is far superior than the global strategy of the ridge and partial least squares solutions.

4.2 Linear Model Setting

For the linear model setting, we adopt the same simulation setup as in Sun et al. (2010), so that the results of our model-free methods are directly comparable to the methods examined in Sun et al. (2010). For self-completeness, we briefly describe the simulation setup here. We first simulate a marker map of 2,000 markers from 20 chromosomes of length 90 cM, with 100 markers per chromosome (using function sim.map in R/qtl (Broman et al., 2003)). The chromosome length is chosen to be close to the average chromosome length in the mouse genome. Then we simulate genotype data of the 360 F2 mice based on the simulated marker map (using function sim.cross in R/qtl). Finally, we choose 10 markers from the 2,000 markers as QTL (Quantitative Trait Locus), and simulate quantitative traits in six cases, as described below, with 100 simulations per case. Given the 10 QTL's, the trait is simulated based on the linear model $y_i=\beta_0+\sum_{j=1}^{p} x_{ij}\beta_j+e_i$, where y_i is the trait value in the ith individual, x_{ij} is the genotype of the jth SNP in the ith individual, which is coded by the number of minor alleles, and e_i is the error term that is normally distributed with mean zero and standard deviation σ_e. Six cases with varying QTL effect sizes are examined:

1. Case 1 – Unlinked QTL's: one QTL per chromosome, with effect sizes 0.5, 0.4, −0.4, 0.3, 0.3, −0.3, 0.2, 0.2, −0.2, and −0.2; $\sigma_e^2=1$.
2. Case 2 – QTL's linked in coupling: two QTL's per chromosome, with effect sizes across five chromosomes as (0.5, 0.3), (−0.4, −0.4), (0.3, 0.3), (0.2, 0.2), and (−0.2, −0.2); $\sigma_e^2=1$.
3. Case 3 – QTL's linked in repulsion: two QTL's per chromosome, with effect sizes across five chromosomes as (0.5, −0.3), (0.4, −0.4), (0.3, −0.3), (0.2, −0.2), and (0.2, −0.2); $\sigma_e^2=1$.

Cases 4, 5, and 6 are the same as cases 1, 2, and 3, respectively, except that $\sigma_e^2=0.5$. The locations and effect sizes of the QTL's in each case are illustrated in Figure C.1 of the Web Appendix C. Moreover, to mimic the reality that the genotype of a QTL may not be observed, we randomly select 1,200 markers as "markers with observed genotype profiles", and only use these 1,200 markers in the multiple loci mapping. There are still high correlations among these 1,200 markers. We estimate the effective number of tests as p_E = 320 (Sun et al., 2010), and set the p-value cutoff at backward filtering as 0.05/320.

An identified marker is defined as a true discovery of a QTL if it is on the same chromosome as the QTL and the R^2 between the marker genotype and the QTL genotype is larger than 0.8. Here 0.8 is chosen so that most unobserved markers can be tagged by an observed marker. We compare variable selection performance of different methods by ROC-like curves that plot the number of true discoveries versus the number of false discoveries across different cutoffs of the coefficient sizes. The ROC-like curve that is close to the

upper-left corner indicates a better performance in that it corresponds to more true discoveries and fewer false discoveries.

Figure 1 reports the results. It is seen from the figure that, overall IAL achieves the best variable selection performance, which is expected, since it is designed for the linear model. mGAP has a slightly inferior but comparable performance as IAL, while it often outperforms elastic net, which in turn outperforms the garrote selection. For instance, for case 4 (unlinked QTL's, σ_e^2=0.5), given zero false discovery, garrote, MAL, elastic net, mGAP and IAL identify 1, 3, 3, 4, 5 true discoveries, respectively. The poor performance of the garrote selection is due to that the signal is rather weak and p is much larger than n, and as a consequence, the initial estimate of the dimension reduction basis is poor, and so is variable selection. Figure 1 depicts the performance of various methods along a range of coefficient cutoffs. In practice, one may often be interested in the case when the cutoff is 0. The corresponding results are shown as the points at the right-upper end of each ROC-like curve in Figure 1. In addition, we also summarize those numbers in Tables D.1 in the Web Appendix D.

4.3 Nonlinear Model Setting

We next consider a number of nonlinear models, each of which is biologically meaningful and is likely to be encountered in real QTL data analysis. For the covariates, we continue to employ the same simulated genotype data as described in the previous section, with the same six cases of varying QTL effect sizes, except that σ_e is now taken as 0.2 and 0.1 respectively.

1. Model 1 – Piecewise linear model: let $x_{\cdot j_k}$ be the j_kth column of X, where j_k is the index of the kth QTL. Let $\tilde{x}=\sum_{k=1}^{9} x_{\cdot j_k} b_{j_k}$, and $\tilde{q}$ = median($\tilde{x}$). Let $y_i = \tilde{x}_i + e_i$ if $\tilde{x} \le \tilde{q}$, and $y_i = \tilde{q} + e_i$ if $\tilde{x} > \tilde{q}$, where $e_i \sim N(0, \sigma_e^2)$. Figure 2(a) illustrates this piecewise linear relation, which, as seen from the plot, is only a mild deviation from the linear model. Such a relation is often seen when the genetic effect is buffered by some other factors.
2. Model 2 – Epistatic interaction: let $\tilde{x}=\sum_{k=1}^{9} x_{\cdot j_k} b_{j_k}$, and $y_i = (\tilde{x}_i + 3)I(x_j \ge 1) + e_i$, where $I(x_{j_{10}} \ge 1)$ is an indicator function. Note the genotype is coded as 0, 1, and 2 for genotype AA, AB, and BB, therefore $I(x_{j_{10}} \ge 1)$ equals to 1 if the genotype is AB or BB, and 0 otherwise. For the particular marker j_{10}, 275 of the 360 samples have genotype AB or BB, and thus the other 9 QTL's only affect the trait in the 275 samples. See Figure 2(b). This type of interaction effect is often observed in genetic studies and is referred to as epistatic interaction (Carlborg and Haley, 2004).
3. Model 3 – Epistatic interaction with unobserved subgroups: let $\tilde{z}$ be an unobserved subgroup indicator, which equals to 0 for 180 samples and 1 for the other 180 samples. Let $\tilde{x}^{(1)} = \Sigma_{k\in\{1,2,7,8,9,10\}} x_{\cdot j_k} b_{j_k}$ and $\tilde{x}^{(2)} = \Sigma_{k\in\{3,4,5,6\}} x_{\cdot j_k} b_{j_k}$. Finally, let $y_i=\tilde{x}_i^{(1)}+e_i$ if $\tilde{z}_i = 0$, and $y_i= -\tilde{x}_i^{(2)}+e_i$ if $\tilde{z}_i = 1$. See Figure 2(c). This is another example of epistatic interaction, but the underlying grouping variable $\tilde{z}$ is unobserved.

Figure 3 shows the results for the epistatic interaction case as depicted in model 2. This model is substantially different from a linear model across all samples, and mGAP is seen to achieve a much better performance than the linear model-based IAL and Elastic Net. MAL often have better variable selection performance than IAL and Elastic Net, but inferior performance than mGAP. For instance, for case 4 (unlinked QTL's and σ_e^2=0.1), given one

false discovery, pSDR, Elastic Net, IAL, MAL and mGAP identifies 1, 2, 2, 3, and 5 true discoveries, respectively. For the sake of space, we report the results of model 1 and model 3 in Figures C.2 and C.3 of the Web Appendix C, respectively. We observe that, the results for model 1 are similar to the linear model case, since the true association is not far from linear, whereas the results for model 3 are similar as the epistatic interaction model 2. The corresponding number of true/false discoveries while coefficient size cutoff is 0 are listed in Supplementary Tables D.2–D.4 in the Web Appendix D.

5. Real Data Analysis

We analyzed the gene expression QTL (eQTL) of 6,100 genes in 112 yeast segregants. The expression of each gene, like other complex traits, is controlled by multiple QTL's (Brem and Kruglyak, 2005), and is often regulated by epistatic interactions between QTL's (Brem et al., 2005). Therefore multiple loci mapping methods, in particular those that can handle epistatic interactions, are important for eQTL studies. In our study, we treated the expressions of those 6,100 genes as separate traits and map their QTL's separately. In other words, our method was evaluated by a comprehensive QTL study on 6,100 traits, which have various levels of heritability and different genetic architectures. We combined the adjacent genetic markers sharing the same genotype profiles. The final data we analyzed consists of 6,100 genes and $p = 1,027$ genotype profiles on $n = 112$ yeast segregants. The effective number of independent tests across these 1,027 genotype profiles was estimated as 412 (Sun and Wright, 2010) and thus the p-value cutoff for backward filtering is set as 0.05/412.

We applied both the linear model-based IAL and the model-free mGAP to this data. mGAP identified 7,594 associations (if one gene is linked to k loci, we counted them as k associations), and at least one QTL for 3,110 (51.0% of 6,100) genes. By contrast, IAL identified 5,262 associations, and at least one QTL for 3,199 (52.4% of 6,100) genes. Table 1 shows the number of genes grouped by the number of QTL's identified by IAL and mGAP. For example, the entry at the first row and the first column is 2,345, which indicates that both IAL and mGAP identify 0 QTL for 2,345 genes. Next we examine this table entry by entry to see which QTL's are captured by one method but missed by the other.

First, 1,047 associations (corresponding to 645 unique genes) are missed by mGAP, but captured by IAL. A closer look reveals that those 1,047 associations have relatively smaller coefficient sizes; see Figure C.4 of the Web Appendix C. Moreover, we conducted a battery of linear model diagnoses for the 645 linear regression models corresponding to those 645 genes, and report the results in Figure C.5 of the Appendix C. It is seen that, for those genes, a linear model provides a reasonable fit. As such, IAL is expected to be more powerful than mGAP, and so identifies QTL's that are missed by mGAP. On the other hand, 1,980 associations (corresponding to 556 unique genes) are missed by IAL, but captured by mGAP. Many of these associations have relatively large effect sizes (Figure C.4), indicating that small effect size is not the major reason that IAL misses those associations. We then carried out diagnosis tests for the 556 models correspond to those 556 genes, and report the results in Figure C.6 of the Appendix C. We found the classical linear model is not sufficient for most of those 556 genes, which again agrees with our observations in simulations that IAL may miss those truly important QTL's due to the deviation from the linear model (e.g., the epistatic effect). In this situation, the model-free mGAP achieves a competent accuracy.

Next, we examine the 1,142 genes with one QTL identified by both IAL and mGAP. The results for the two methods are consistent for those genes, where the locations of the QTL's from IAL and mGAP are within 50kb for 1,074 (94%) genes.

There 420 cases that IAL identifies one QTL for a gene, but mGAP identifies two QTL's for this gene, and vice versa (Table 1, cells highlighted by bold typeface). It is often the case that the two QTL's identified by one method includes the single QTL identified by the other method (about 92% with distance < 50kb). The difference is that the QTL pairs that are identified only by mGAP capture epistatic interactions more frequently than the QTL paris that are identified only by IAL; see Figure 4 of the paper.

For the remaining 992 genes in Table 1 that are highlighted by underscore, similar conclusions can be drawn that a larger proportion of models identified by mGAP does not fit a linear model well (Figures C.7 and C.8 of the Web Appendix C). For example, about 19% (192 out of 992) of the models identified by IAL show heteroscedasticity (p-value < 0.05), in contrast to ~31% (311 out of 992) heteroscedastic models identified by mGAP. Among the 384 genes where IAL identifies two QTL's, ~13% (49 out of 384) show significant two-way interaction (p-value < 0.05); while among the 313 genes where mGAP identifies two QTL's, ~19% (60 out of 313) show significant two-way interaction.

6. Discussions

Despite the flourish of proposals on variable selection for high-dimensional data in recent years, genome-wide multiple loci mapping has remained to be challenging. This can be attributed to its high dimensionality but a relatively weak signal, as well as the presence of complex associations such as epistatic interactions. In this article, we have proposed and examined a number of model-free variable selection solutions for small-*n*-large-*p* regressions. Compared with the existing model-based solutions, our proposal of multivariate group-wise adaptive penalization (mGAP) retains a comparable accuracy when the imposed model and the true one agree, whereas it achieves a far superior performance when the underlying model deviates from the assumed one. This is particularly useful, since, as illustrated in our real data studies across 6,100 gene expression traits, the true association in QTL mapping is often more complex than the usually imposed linear model, and is generally unknown *a priori*. In practice, our proposed method can be used in conjunction with, rather than as an alternative to, many model-based variable selection approaches. One example is given in our analysis of yeast segregants data, where the best practical strategy is a joint application of both the model-based IAL and the model-free mGAP.

Within the model-free variable selection framework, we have also compared the global strategy of the ridge solution and the PLS solution with the local strategy of the mGAP solution. The key difference is that the former handles all p covariates simultaneously, whereas the latter takes an iterative approach by updating one variable at a time. Another distinction is that the ridge and PLS solutions perform simultaneous dimension reduction basis estimation and variable selection, while mGAP focuses on variable selection only. We present the results of the global strategy since both ridge and PLS are natural extensions of the existing solutions from $n > p$ to $n < p$. Based on our simulation studies, we believe the local strategy is a more effective solution for high dimensional data.

We have implemented mGAP, ridge, PLS methods, as well as modified version of MAL and elastic net in an R package BPrimm (Bayesian and Penalized regression in multiple loci mapping), which can be freely downloaded from http://www.bios.unc.edu/~wsun/software/BPrimm.htm. The computational intensive part of mGAP is implemented by C. The computational time and memory load of mGAP is reasonable for large scale QTL mapping. For example, in our real data studies across 6,100 gene expression traits, with $n = 112$ and $p = 1, 027$, mGAP takes less than 1G memory in total and about 50 seconds for multiple loci mapping of each trait. For studies with a large

number of traits, parallel computation with a small number of traits per CPU is recommended.

There are a number of avenues for future extensions. First, we have concentrated on the genome-wide multiple loci mapping in experimental cross of inbred strains, while similar problems exist for genome-wide association studies (GWAS) in human population. Genetic markers in experimental cross tend to have higher correlations than in GWAS. Therefore, simultaneous multiple loci selection is expected to be more advantageous than marginal regression in experimental cross than in GWAS. Nevertheless it is of interest to consider the GWAS data. Second, the mGAP solution has been designed to handle a univariate response, i.e., a single trait at a time. In order to perform mGAP across a large number of gene expression traits, we have chosen to model each trait separately in our real data analysis, and the multiple testing across gene expression traits is not corrected for. It is an intriguing question to study multiple traits simultaneously so to borrow information across the correlated traits. These questions are currently under investigation.

7. Supplementary Materials

The Web Appendix, Tables, and Figures referenced in Sections 3,4,5 are available under the Paper Information link at the *Biometrics* website http://www.biometrics.tibs.org.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sun's research is supported in part by NIH RC2 MH089951-01, P50-MH090338-01 and EPA RD-83382501. Li's research is supported in part by NSF grant DMS 0706919. The authors thank the editor, the associate editor, and the two referees for their valuable suggestions.

References

Bondell HD, Li L. Shrinkage inverse regression estimation for model-free variable selection. Journal of The Royal Statistical Society Series B. 2009; 71:287–299.

Breiman L. Better subset regression using the nonnegative garrote. Technometrics. 1995; 37:373–384.

Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A. 2005; 102:1572–1577. [PubMed: 15659551]

Brem RB, Storey JD, Whittle J, Kruglyak L. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature. 2005; 436:701–703. [PubMed: 16079846]

Broman K, Wu H, Sen S, Churchill G. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003; 19:889–890. [PubMed: 12724300]

Candés EJ, Tao T. The dantzig selector: statistical estimation when *p* is much larger than *n*. The Annals of Statistics. 2007; 35:2313–2351.

Carlborg Ö, Haley C. Epistasis: too often neglected in complex trait studies? Nature Reviews Genetics. 2004; 5:618–625.

Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika. 2008; 95:759.

Cook RD. Graphics for regressions with a binary response. Journal of the American Statistical Association. 1996; 91:983–992.

Cook, RD. Regression Graphics: Ideas for Studying Regressions Through Graphics. Wiley; 1998.

Cook RD. Testing predictor contributions in sufficient dimension reduction. Annals of Statistics. 2004; 32:1062–1092.

Cook RD, Ni L. Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. Journal of the American Statistical Association. 2005; 100:410–428.

Cook RD, Ni L. Using intraslice covariances for improved estimation of the central subspace in regression. Biometrika. 2006; 93:65–74.

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association. 2001; 96:1348–1360.

Fan J, Lv J. Sure independence screening for ultra-high dimensional feature space. Journal of the Royal Statistical Society Series B. 2008; 70:849–911.

Fung W, He X, Liu L, Shi P. Dimension reduction based on canonical correlation. Statistica Sinica. 2002; 12:1093–1114.

Hall P, Li KC. On almost linearity of low dimensional projections from high dimensional data. Annals of Statistics. 1993; 21:867–889.

Hoggart C, Whittaker J, De Iorio M, Balding D. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. PLoS Genetics. 2008:4.

Li KC. Sliced inverse regression for dimension reduction. Journal of the American Statistical Association. 1991; 86:316–327.

Li L, Cook RD, Tsai CL. Partial inverse regression. Biometrika. 2007; 94:615–626.

Li L, Dennis Cook R, Nachtsheim C. Model-free variable selection. Journal of the Royal Statistical Society Series B(Statistical Methodology). 2005; 67:285–299.

Li L, Yin X. Rejoinder to “A note on sliced inverse regression with regularizations”. Biometrics. 2008a; 64:984–986.

Li L, Yin X. Sliced inverse regression with regularizations. Biometrics. 2008b; 64:124–131. [PubMed: 17651455]

Ni L, Cook RD, Tsai CL. A note on shrinkage sliced inverse regression. Biometrika. 2005; 92:242.

Ni L, Wang H, Tsai CH, Zhou J. Model Free Variable Selection via Adaptive Lasso. Proceedings of the Joint Statistical Meetings. 2008

Park, T.; Casella, G. Technical report. University of Florida; Gainesville, FL: 2008. The Bayesian Lasso.

Sun W, Ibrahim J, Zou F. Genomewide Multiple-Loci Mapping in Experimental Crosses by Iterative Adaptive Penalized Regression. Genetics. 2010; 185:349. [PubMed: 20157003]

Sun W, Wright FA. A geometric interpretation of the permutation p-value and its application in eQTL studies. Annals of Applied Statistics. 2010; 4:1014–1033.

Tibshirani R. Regression shrinkage and selection via the Lasso. J Royal Statist Soc B. 1996; 58:267–288.

Wang H. Forward regression for ultra-high dimensional variable screening. Journal of the American Statistical Association. 2009; 104:1512–1524.

Wold H. Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. Perspectives in Probability and Statistics (papers in honour of MS Bartlett on the occasion of his 65th birthday). 1975:117–142.

Yi N. A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. Genetics. 2004; 167:967–975. [PubMed: 15238545]

Yi N, Xu S. Bayesian LASSO for Quantitative Trait Loci Mapping. Genetics. 2008; 179:1045–1055. [PubMed: 18505874]

Yin X, Cook RD. Dimension reduction for the conditional kth moment in regression. Journal of the Royal Statistical Society, Series B. 2002; 64:159–175.

Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. Journal Of The Royal Statistical Society Series B. 2006; 68:49–67.

Yuan M, Lin Y. On the nonnegative garrote estimator. Journal of the Royal Statistical Society Series B. 2007; 69:143–161.

Zellner A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association. 1962; 57:348–368.

Zhou J, He X. Dimension reduction based on constrained canonical correlation and variable filtering. Annals of Statistics. 2008; 36:1649–1668.

Zou H. The adaptive Lasso and its oracle properties. Journal of the American Statistical Association. 2006; 101:1418–1429.

Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B. 2005; 67:301–320.

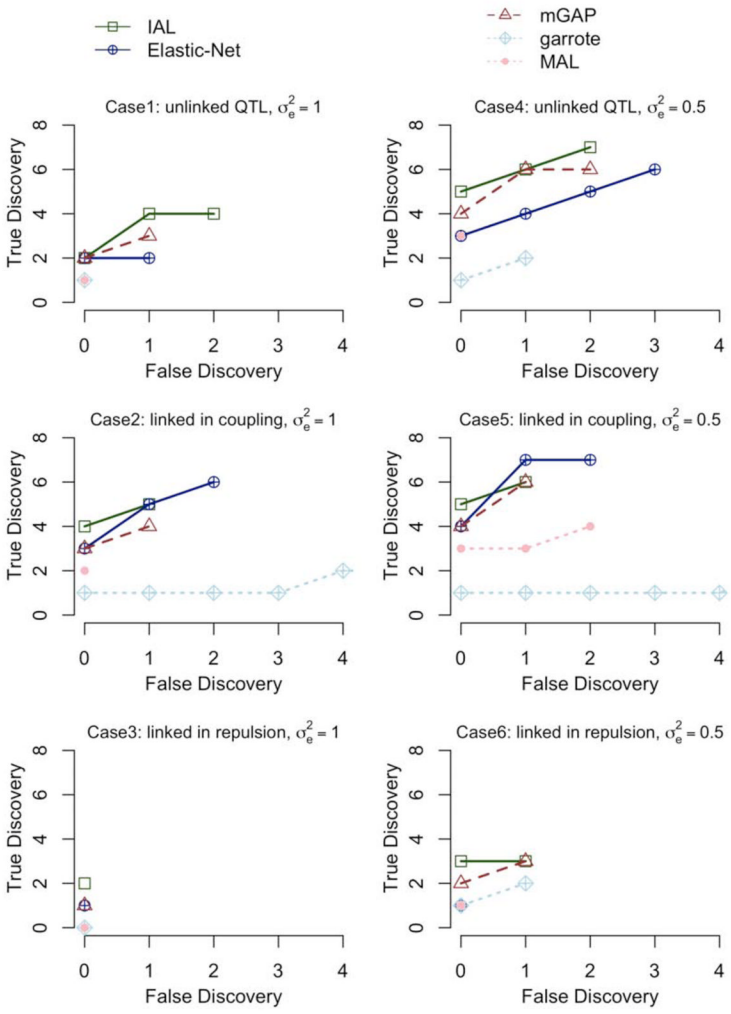

Figure 1.
Comparison of IAL, Elastic Net, mGAP, pSDR (PLS solution), and MAL (PLS initial estimates) under the linear model setting.

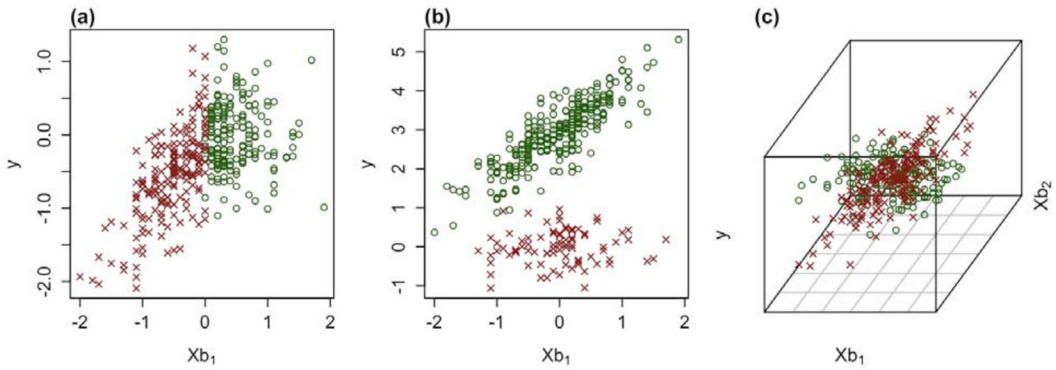

Figure 2.
Illustration of the three simulated nonlinear models. Two types of points indicate two subgroups of the samples. Xb_1 and Xb_2 indicate two linear combinations of the covariates.

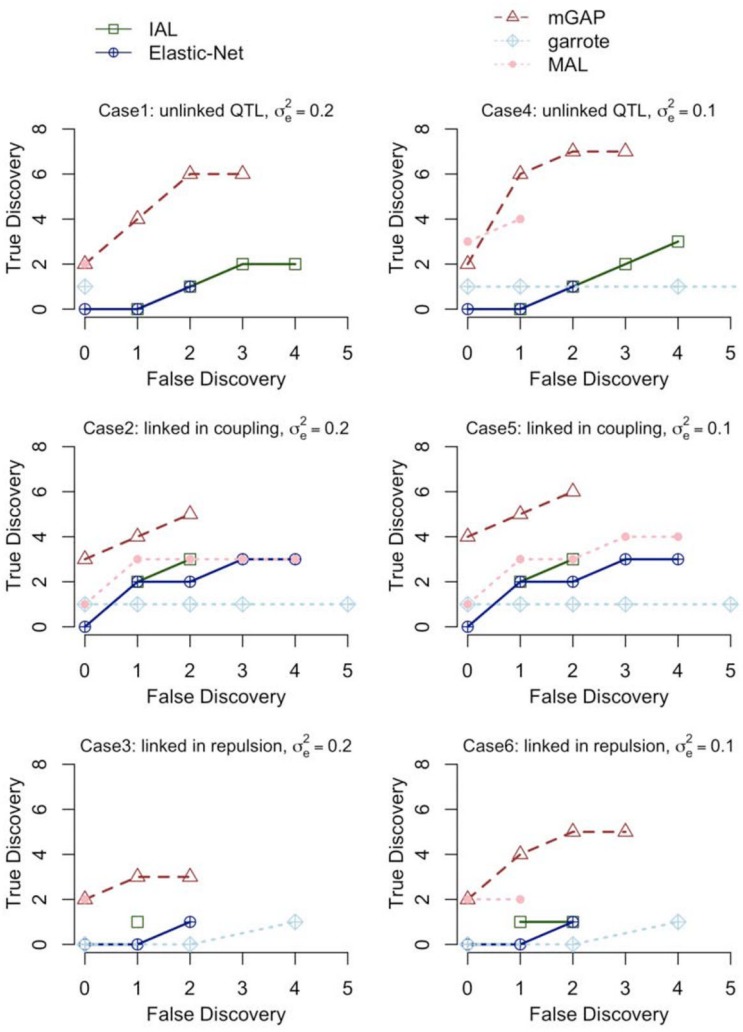

Figure 3.
Comparison of IAL, Elastic Net, mGAP, pSDR (PLS solution), and MAL (PLS initial estiamtes) under the nonlinear model setting when there are epistatic interactions.

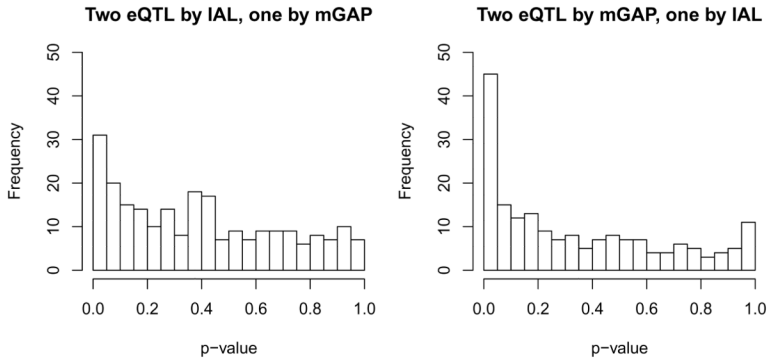

Figure 4.

Comparison of IAL and mGAP in terms of interaction p-values, which are calculated by ANOVA test comparing the additive model and the model with both additive and interaction effects. Some genes are linked to two QTL's by one method but one eQTL by the other method. Here we show that for these genes, the QTL pairs identified by the mGAP (right panel, 185 genes) are more likely to capture the interaction effect than the QTL pairs identified by the IAL (left panel, 235 genes) since there are more small p-values in the right panel than in the left penal.

Table 1

The number of genes with certain number of QTL's identified by mGAP and IAL. For example, the entry at first row and first column is 2345, which indicates that both IAL and mGAP identify 0 QTL for 2,345 genes.

		# of QTL's by mGAP			
		0	**1**	**2**	**≥ 3**
# of QTL's by IAL	0	2345	224	104	228
	1	398	1142	**185**	209
	2	152	**235**	225	159
	≥3	95	147	88	164